

# **Zusammenfassung DASB**

## Data Science Basics

Maurin D. Thalmann

23. Januar 2020

# Inhaltsverzeichnis

<b>1</b>	<b>Introduction to Data Science</b>	<b>2</b>
1.1	Introductory Stizzle . . . . .	2
1.2	A Global View on Data Science . . . . .	2
1.3	Different Concepts in Data Analytics . . . . .	3
1.3.1	1993 - Online Analytical Processing (OLAP) . . . . .	3
1.3.2	1998 - Fuzzy Data Analysis . . . . .	3
1.3.3	2005 - Data Mining . . . . .	4
1.3.4	2013 - Predictive Modeling . . . . .	4
1.3.5	2013 - Data Driven Decision Making . . . . .	4

# 1 Introduction to Data Science

## 1.1 Introductory Stizzle

- More is different - with different quantity comes different quality, although „more“ is always dependent on the point of view.
- Data volume nowadays is exploding (276.12 billion GB of digital data)
- Data Science tries to close the gap between „Data available to an organization“ and „Percent of data an organization can process“.

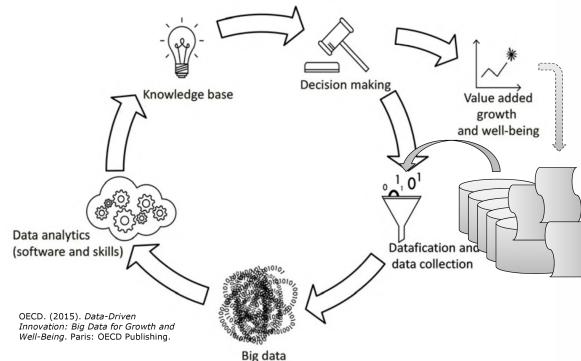


Abbildung 1: Beautiful illustration of the classic data value cycle

„**Data Science** is the extraction of actionable knowledge directly from data through a discovery, or hypothesis formulation and hypothesis testing“. Data Scientists generate knowledge from big data.

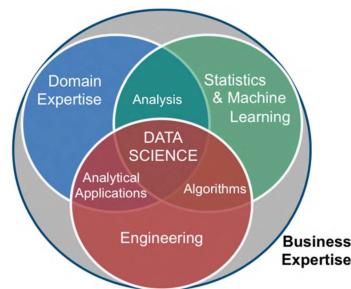


Abbildung 2: Skills needed in Data Science

## 1.2 A Global View on Data Science

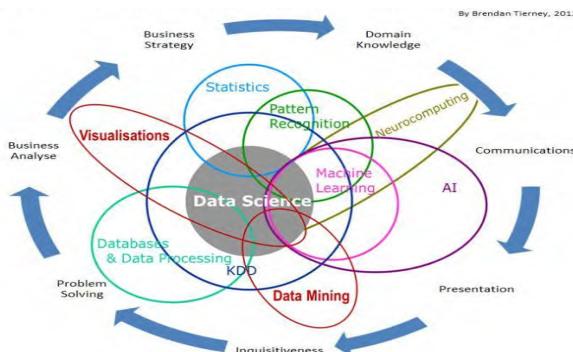


Abbildung 3: Data Science is multidisciplinary

- **AI**: programs that perform tasks resembling humans (learn and reason)
- **ML**: algorithms to learn from that data without explicit programming
- **DL**: subset of ML using artificial neural networks for treating vast amount of data (big data)
- **Data Science**: spans the collection, management, analysis and interpretation of large amounts of data with a wide range of applications → make informed decisions based on what was learned
- **EDA** (exploratory data analysis): extract insight from data (outside AI, human based)

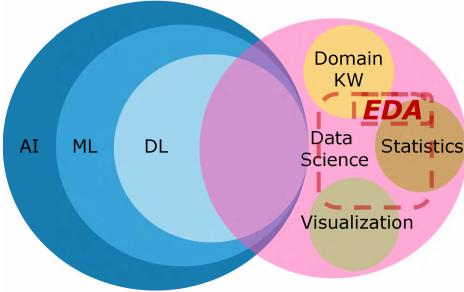
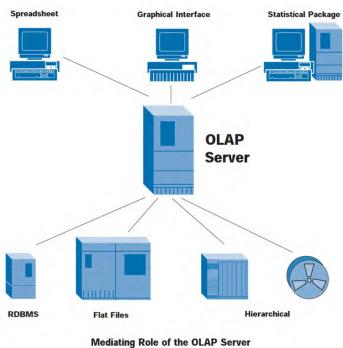


Abbildung 4: AI / ML / DL and Data Science

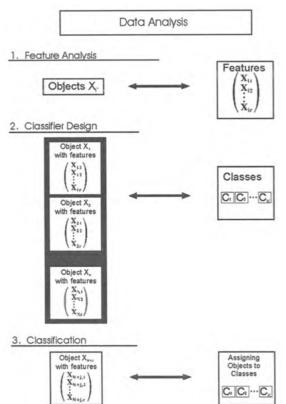
### 1.3 Different Concepts in Data Analytics

#### 1.3.1 1993 - Online Analytical Processing (OLAP)



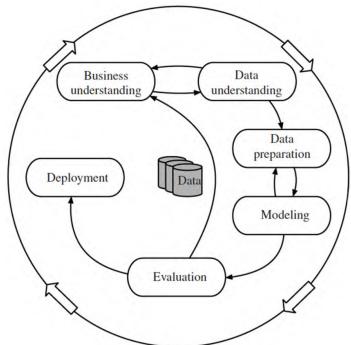
- It's online, not batch (interactive, not programmed in COBOL)
- The OLAP system should access the data required to perform the indicated analysis
- OLAP tools empower useranalysts to easily perform multi-dimensional analysis, which previously have been avoided because of their perceived complexity.

#### 1.3.2 1998 - Fuzzy Data Analysis



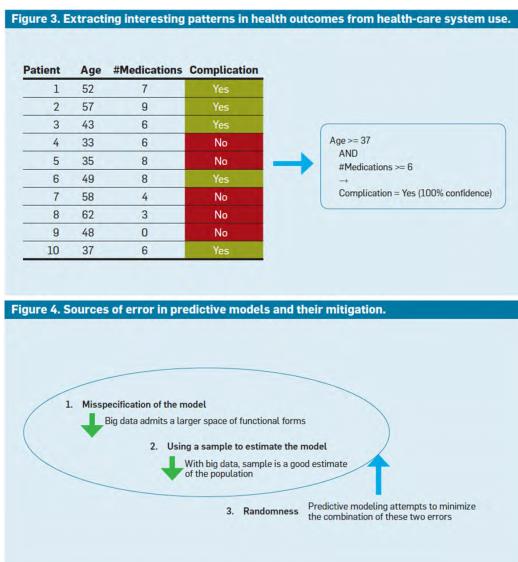
- Data analysis can be defined as **search for structure** in data
- In data analysis, objects are considered which are described by some attributes
- Most of the traditional methods for data analysis assume that patterns to be detected are two-valued
- Whenever this is not the case, the relationship between data and classes becomes gradual  
→ Fuzzy Classification

### 1.3.3 2005 - Data Mining



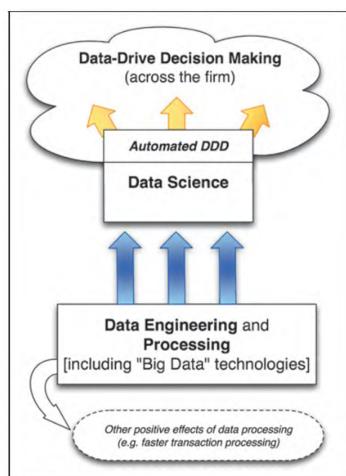
- Very similar concept to data science
- Machine Learning (modeling) is the technical core of practical data mining applications
- Data Mining is a business process related to *value* (finding the metaphorical gold nugget)
- The lifecycle of a data mining project is defined by the CRISP-DM reference model

### 1.3.4 2013 - Predictive Modeling



- A common epistemic requirement in assessing whether new knowledge is actionable for decision making in its *predictive power*, not just its ability to explain the past.
- The requirement on predictive accuracy on observations that will occur in the future is a key consideration in data science.

### 1.3.5 2013 - Data Driven Decision Making



- Data science involves much more than just data-mining algorithms.
- Successful data scientists must be able to view business problems from a *data perspective*.

## **2 Intro to R and Exploratory Data Analysis (EDA)**

### **2.1 Intro to R**